Reasoning with uncertainty: intro

Teadmiste esitamise kursuses

Uncertainties everywhere

Almost all the knowledge we have is uncertain: there are many exceptions to a rule or a fact/rule holds with some vague probability.

Notice that, in contrast, typical databases in companies contain facts with very high certainty, but they do not contain rule or commonsense knowledge: the latter is built into brittle software code of systems using these databases.

In contrast, mathematics is certain, because we define it

Discrete and numeric methods

In practice there are two main different ways of tackling uncertainty:

- Numeric methods: uncertainty is estimated with probability-like numbers of various kinds.
- Discrete methods: no numbers used. Instead, uncertainty is described as exceptions or beliefs of people.

Discrete example from default logic: birds

bird(tweety).

penguin(pennie).

penguin(X) => bird(X).

penguin(X) => -flies(X).

Use (bird(X) => flies(X)) if we cannot prove that –flies(X).

Query: flies(X) ?

Common syntax for default rules:

$$\frac{\text{bird}(X)\ :\ \text{flies}(X)}{\text{flies}(X)}$$

Another classic example from default logic

republican(nixon).

quaker(nixon).

Use (republican(X) => -pacifist(X)) if we cannot prove that pacifist(X).

Use (quaker(X) => pacifist(X)) if we cannot prove that -pacifist(X).

Query: pacifist(nixon) ?

Some initial notes about default logic

- Similar to the somewhat vague concept of “defeasible logic”
- Since first order logic (FOL) is undecidable, it is generally impossible to check that a rule can be fired!
- Practical systems are almost all focused on propositional instances of FOL formulas: propositional logic is decidable, thus rules can be checked.
- Nonmonotonic logic: all standard logics are monotonic, default logic is not

Numeric methods: a motivating example

John observes something in the field which looks like a bird and estimates the probability of it being a bird as 80%. Mike observes the same object, but estimates the probability as 70%. I have an intuitive rule which says that birds can fly with the probability 90%. Let us formalize this as:

bird(object1) : 0.8 :: John
bird(object1) : 0.7 :: Mike
[bird(X) => canfly(X)] : 0.9 :: me

What can I derive from here? A simple idea is to combine the bird observation probabilities to a stronger one, using a standard probability calculation rule P(a v b) = P(a) + P(b) - (P(a) * P(b)), which holds in case a and b are independent observations. We get:

bird(object1) : 0.8+0.7-(0.9*0.7) = 0.87 :: John, Mike

Second, using the derivation rule, decreasing the probability according to a probability rule P(a & b) = P(a) * P(b), again in case when a and b are independent:

canfly(object1) : 0.87 * 0.9 = 0.783 :: John,Mike,me

What can go wrong in this example?

We have to be careful to indicate what the probability number means: it can be interpreted in several ways. For example, does

bird(object1) : 0.1

mean that object1 is unlikely to be a bird, or that it is a bit likely to be a bird? In other words, are small probability numbers really interpreted as positive or rather indications of a probability of a negation

-bird(object1) : 0.9

What more could go wrong or require extra care?

- Are the calculation rules we used really true or applicable? In which context are they applicable? Can they lead to incorrect or unintuitive results in other cases?
- What about "independent": are these statements independent and to what degree?
- Are we sure our procedure does not use same input several times to "over-strengthen" the probabilities?
- What is the "formalism" we used, ie how to encode that a rule/fact has a probability and sources? Can we use quantifiers also there, like quantifying over sources or something?

Different types of uncertainties

There are many different camps of people advocating different ways to understand and handle uncertainty. The main camps are:

- Frequentists: probabilities as limit values of statistical experiments, when more and more experiments are made. Dice-kinds of examples fit this camp very well. Also called physical interpretation.
- Bayesianists: more subjective, interpreting probability as "reasonable expectation" (a probability is assigned to a hypothesis), yet basing their calculations on the famous Bayes theorem. Also called evidential interp.

Bayesianists are similar to

- Subjectivists: probability measures a "personal belief“ (think making bets)

Main mathematical grounds

Frequentists base probability theory on the three axioms from Kolmogorov.

A simplified version of these axioms:

- **Axiom 1:** The probability of an event is a real number greater than or equal to 0.
- **Axiom 2:** The probability that at least one of all the possible outcomes of a process (such as rolling a die) will occur is 1.
- **Axiom 3:** If two events *A* and *B* are mutually exclusive, then the probability of either *A* or *B* occurring is the probability of *A* occurring plus the probability of *B* occurring.

See https://en.wikipedia.org/wiki/Probability_axioms for the real ones

Bayes rule

Allows you to find $P(A|B)$ from $P(B|A)$, i.e. to 'invert' conditional probabilities.

$$P(A|B) = [P(B|A) \cdot P(A)] \;/\; P(B)$$

Often compute the denominator $P(B)$ using the law of total probability.

Bayes' Rule - Updating Probabilities

- Let $A_1,\ldots,A_k$ be a set of events that **partition** a sample space such that (mutually exclusive and exhaustive):
  - each set has known $P(A_i) > 0$ (each event can occur)
  - for any 2 sets A_i and A_j, $P(A_i \text{ and } A_j) = 0$ (events are disjoint)
  - $P(A_1) + \ldots + P(A_k) = 1$ (each outcome belongs to one of events)
- If C is an event such that
  - $0 < P(C) < 1$ (C can occur, but will not necessarily occur)
  - We know the probability will occur given each event A_i: $P(C|A_i)$
- Then we can compute probability of A_i given C occurred:

$$P(A_i \mid C) = \frac{P(C \mid A_i)P(A_i)}{P(C \mid A_1)P(A_1) + \cdots + P(C \mid A_k)P(A_k)} = \frac{P(A_i \text{ and } C)}{P(C)}$$

Example - OJ Simpson Trial

- Given Information on Blood Test (T+/T-)
  - Sensitivity: P(T+|Guilty)=1
  - Specificity: P(T-|Innocent)=.9957 → P(T+|I)=.0043
- Suppose you have a prior belief of guilt: P(G)=p*
- What is "posterior" probability of guilt after seeing evidence that blood matches: P(G|T+)?

$$P(T+) = P(T^+G) + P(T^+I) = P(G)P(T^+ \mid G) + P(I)P(T^+ \mid I) =$$

$$= p^*(1) + (1 - p^*)(.0043)$$

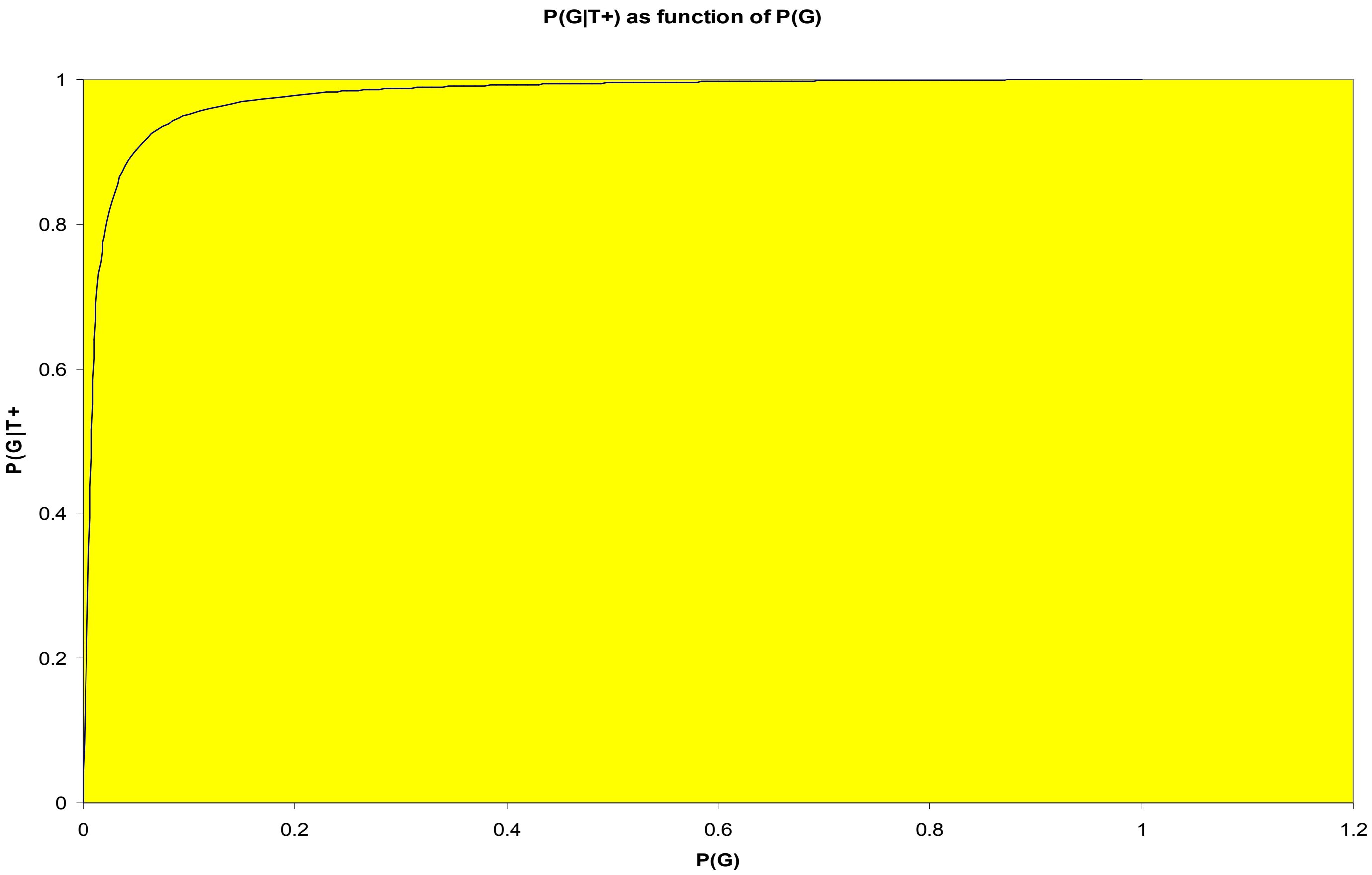
$$P(G \mid T^+) = \frac{P(T^+G)}{P(T^+)} = \frac{P(G)P(T^+ \mid G)}{P(T^+)} = \frac{p^*(1)}{p^*(1) + (1 - p^*)(.0043)} = \frac{p^*}{.9957p^* + .0043}$$

Source: B.Forst (1996). "Evidence, Probabilities and Legal Standards for Determination of Guilt: Beyond the OJ Trial", in *Representing OJ: Murder, Criminal Justice, and the Mass Culture*, ed. G. Barak pp. 22-28. Harrow and Heston, Guilderland, NY

OJ Simpson Posterior (to Positive Test) Probabilities

$$\text{Prior Probability of Guilt}: \qquad P(G) = .10 \Rightarrow$$

$$P(G \mid T^{+}) = \frac{.10(1)}{.10(1) + .90(.0043)} = \frac{.10}{.10387} = .9627$$



Different measures of uncertainty

Broadly said, the main two measures are:

- Standard: red(a1): p indicates that the probability of a1 being red is p.
- Fuzzy: red(a1): p indicates how red a1 is: (1.0 means totally red)

Fuzzy measures lead to fuzzy logic, which is a popular field.

Intervals of probability

It may happen that we know the lower and upper limits of probability, but not the exact probability.

Fits well for modeling sensors which can sometimes give erroneous values, while we know – from stats – their error rates.

The most famous approach in this direction is the Dempster-Schafer theory of evidence.